MAPPING GRAZED VEGETATION COMMUNITIES ON MACQUARIE ISLAND USING A BINARY ENSEMBLE CLASSIFIER

Arko Lucieer
University of Tasmania
School of Geography and Environmental Studies
Private Bag 76
Hobart, Tasmania 7001
Australia
Ph. +61 362262140, Fax. +61 362267628
Email: Arko.Lucieer@utas.edu.au
Web: http://www.lucieer.net

Abstract

This study implemented and applied a binary ensemble classifier for identification of grazed vegetation communities on Macquarie Island from very high resolution Quickbird imagery. Rabbit grazing has severely affected Macquarie's unique sub-Antarctic vegetation communities. The aim of this study was to identify the grazed areas from Quickbird imagery to map their spatial extent. Seven different soft classification algorithms were applied to classify the image into grazed vs. 'other' classes. The maximum likelihood classifier, supervised fuzzy *c*-means classifier (Euclidean distance, Mahalanobis distance, and *k*-nearest neighbour), and three support vector machine classifiers (SVM) were applied. An ensemble classifier based on the consensus rule was used to combine the seven classification results. A very high classification accuracy of 97% was achieved with the ensemble classifier, identifying grazed areas and providing an estimate of classification uncertainty.

Introduction

Image classification based on satellite imagery is a widely used technique for extracting thematic information on land cover. We classify image pixels by reducing spectral information in multiple bands to a relatively small number of general classes (Tso and Mather, 2001). There is a wide range of image classification algorithms that can be applied to remote sensing classification problems. A recent review of classification algorithms (Lu and Weng, 2007) shows the large number of classifiers available to the remote sensing analyst. A grouping of classifiers into supervised, unsupervised, soft, hard, parametric, non-parametric, contextual, object-based algorithms indicates that each of these techniques makes different assumptions about the training samples, statistical distributions, class separability, and decision boundaries. With an accuracy assessment the performance of classifiers can be tested and the best classification result selected, however, because of the different classification approaches each classification result can contain valuable information that is underutilised if only one classification result is used. A recent development in remote sensing classification is the use of multisource or ensemble classifiers (Benediktsson, 1999; Benediktsson et al., 2007; Doan and Foody, 2007; Tzeng et al., 2008). The ensemble-based approach is a multiple classifier system in which the aim is to combine the outputs of several classifiers to derive an accurate classification (Foody et al., 2007).

Most land cover classifications aim to map all classes in the image or study area. This means that training information is required for all classes. In a lot of studies, however, the focus might only be on one or two key classes, e.g. studies of invasive species, change detection of impervious areas, or detection of forest logging. For these types of applications we might be able to focus on a single class as the other classes might be of no interest. This single class or binary classification approach has the advantage that it reduces the requirements for the collection of training samples. In addition, a binary classification could improve the separation in feature space between the focus class and the class 'other' (Boyd et al., 2006).

In this study, I follow the ensemble binary classification approach taken by Foody (2007) in mapping areas affected by rabbit grazing on sub-Antarctic Macquarie Island from very high resolution (VHR) imagery. The aim is to accurately map the vegetated areas that have been severely grazed by rabbits by applying a binary classification with different classifiers and combining the results in an ensemble-classification approach.

Study area and imagery: Macquarie Island

Macquarie Island is a remote sub-Antarctic island approximately equidistant between Tasmania, New Zealand and Antarctica in the Southern Ocean (location: 54º S 159º E; approximate size: 35km by 5km). The Macquarie Island Nature Reserve is one of the most valuable reserves in Australia and the World, well recognised for its conservation, geological, ecological and scientific values. It is a World Heritage Area, a Biosphere Reserve, and listed on the Register of the National Estate. Rabbits were introduced in the 1870s and have had major impacts on most of the reserve's plant species. Heavy rabbit grazing has resulted in the destruction of tall tussock grassland and Macquarie Island cabbage which, in turn, has had a devastating impact on the population of many burrowing sea bird species through habitat destruction (Bergstrom and Chown, 1999; Copson and Whinam, 1998; Kirkpatrick and Scott, 2002). Up-to-date and accurate spatial data, such as vegetation maps, are of crucial importance for sustainable management of the island. Because of the island's remoteness, satellite imagery provides advanced, efficient, and non-invasive means to map its land cover and to quantify environmental changes.

For this study, a cloud-free Quickbird image of Macquarie Island acquired on 18 March 2007 is used for land cover classification. DigitalGlobe's Quickbird satellite captures four multispectral bands (Blue, Green, Red, and Near-Infrared (NIR)) at 2.4 m resolution and one panchromatic band covering the visible and NIR parts of the electromagnetic spectrum at 0.6 m resolution (DigitalGlobe, 2008). Fig. 1 shows a colour composite of the Red, Green, and Blue bands (visible) of the Quickbird image; the image subset of the study area, used to illustrate the classification techniques proposed in this study, is shown as a red box. The image was orthorectified using 26 survey marks spread out over the island and a 5 m DEM acquired by NASA's AIRSAR in 2002.

The aim of the classification is to map vegetation communities that are indicative for rabbit grazing. Tall tussock grassland dominated by *Stilbocarpa Polaris* (Macquarie Island cabbage) and *Poa foliosa* is characteristic of Macquarie Island's steep coastal slopes. These plant species are very palatable to the rabbits on the island and as a result large areas of the coastal slopes have been grazed and degraded. When the Tussock plants have been grazed they turn into a yellow-golden colour and eventually die off, exposing soil and resulting in erosion. Some of these bare areas are later recolonised by short grassland or *Acaena* monostands replacing the original vegetation. The grazed areas have a very characteristic spectral signature with a high overall reflectance and a drop in red absorption and near-infrared reflectance. In this natural area there are no hard boundaries between the vegetation communities and land cover classes. Transition zones (ecotones) between the vegetation communities and between grazed and non-grazed areas are characteristic for the island. A classification approach based on soft image classification is therefore most appropriate in order to quantify the uncertainty related to ecotones.

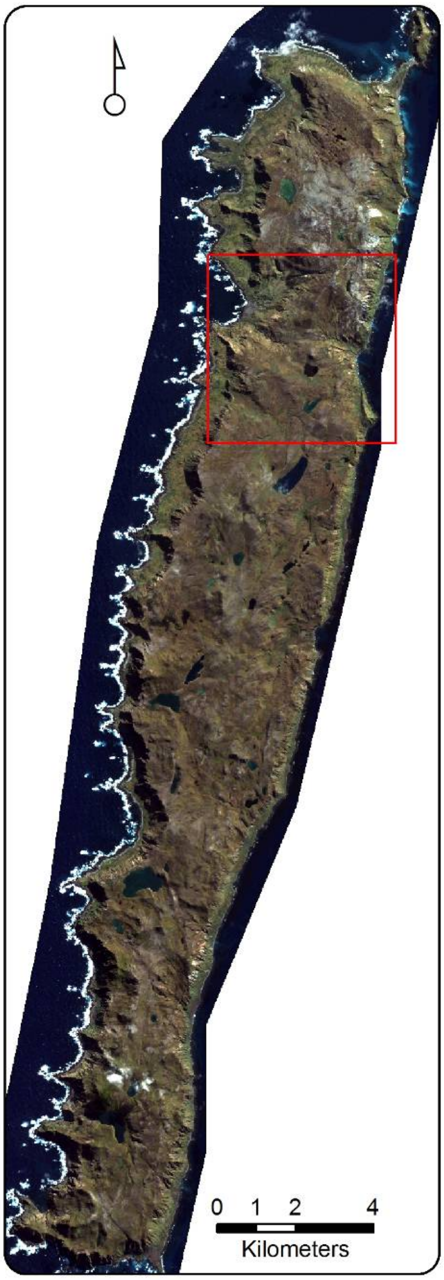


Fig. 1. Colour composite of the visible multispectral bands of the Quickbird image of Macquarie Island, one of the unique sub-Antarctic islands in the Southern Ocean. The study area is highlighted by a red square.

Methods

Classification

Five soft classification algorithms are applied in this study. The advantage of soft classifiers is that they give a measure of classification likelihood for each pixel to each class in the form of probabilities or fuzzy membership values. This study applied the following classification algorithms.

1. Maximum likelihood: parametric
2. Supervised fuzzy *c*-means (SFCM) Euclidean distance: parametric
3. SFCM Mahalanobis distance: parametric

4. SFCM *k*-nearest neighbour (*k*NN): non-parametric
5. Support Vector Machine (SVM): non-parametric

The maximum likelihood classifier is a widely used classification algorithm based on Bayes' theorem and founded in probability theory. This classifier is parametric in the sense that it models the statistical distribution of classes in multivariate feature space by class means and variance-covariance matrices, effectively approximating the class shape with a hyper-ellipsoid. The main (sometimes limiting) assumption of this classifier is that it assumes a multivariate normal distribution of spectral values of pixels within a class (Lu and Weng, 2007; Richards and Jia, 2005; Tso and Mather, 2001). The maximum likelihood classifier is here considered a soft classifier as it provides probabilities for each class for each pixel, giving an indication of the strength of reliability of the classification.

Fuzzy classification is based on the concept of fuzzy sets (Zadeh, 1965). In the fuzzy set model, the class assignment function attributes to each element a grade of membership in the real interval [0,1] for every defined set. This grade of membership corresponds to the degree to which the element (i.e. pixel) is similar to the concept or prototype represented by that set (i.e. land cover class). The well-known unsupervised fuzzy *c*-means classifier (FCM) uses an iterative procedure that starts with an initial random allocation of pixels to be classified into *c* clusters. Given the cluster allocation, the centre of each cluster is calculated as the weighted average of the pixel spectral values. In the next step, pixels are reallocated among the classes according to the relative similarity between pixels and clusters based on a well-known distance measure, such as the Euclidean or Mahalanobis (both variance and covariance are used for distance scaling) metrics. Reallocation proceeds by iteration until a stable solution is reached where similar pixels are grouped together in a cluster. Their membership value gives their degree of affinity with the centroid of the class (Bezdek, 1981). The fuzzy *c*-means clustering algorithm (FCM) is unsupervised meaning that the resulting class clusters are unlabeled. In most remote sensing applications, however, a supervised approach is more common where an expert can train the classifier by selecting clusters of reference pixels in the image that represent land cover classes. In this study, a modified version of the fuzzy *c*-means algorithm, based on Zhang and Foody (2001) is applied in order to develop a *supervised* fuzzy classification method. In the supervised fuzzy *c*-means, the class centroids are determined from the mean of the training pixels. This reduces the clustering algorithm to a one step calculation, resulting in fuzzy membership values for each pixel in each of the defined classes.

The similarity of a pixel to a class is expressed by a membership value μ, which is defined as

$$\mu_{ic} = \frac{(d_{ic}^2)^{-\left(\frac{1}{q-1}\right)}}{\sum_{c=1}^{k}(d_{ic}^2)^{-\left(\frac{1}{q-1}\right)}} \tag{1}$$

where μ_{ic} is the membership value of the i^{th} sample to class c, d_{ic} is the distance between the sample i and cluster centre c in feature space, k is the number of clusters and q is the fuzziness exponent representing the degree of class overlap. Generally the Euclidean distance (d_{ic}) from a sample vector (g_i) to a cluster mean (m_c) is taken to be the similarity criterion, $\|g_i - m_c\|$. The degree to which a sample belongs to a class is expressed not in terms of a binary 'yes' or 'no' but by a continuous membership value that ranges between 0.0 and 1.0, where 1.0 indicates perfect similarity with the cluster centroid. The parameter q ($q > 1.0$) is the fuzzy exponent, which determines the amount of overlap in the cluster model. Values between 1.5 and 3.0 are commonly found in literature, but a value of 2.0 is most widely used (Burrough et al., 2000; Foody, 1996).

Three different supervised fuzzy *c*-means (SFCM) classifiers were implemented and applied in this study. The first uses the Euclidean distance as the distance metric, i.e. only the class mean is used which makes this algorithm the fuzzy equivalent of the minimum distance to mean algorithm. The second algorithm uses the Mahalanobis distance metric based on the class mean and covariance matrix, modelling the classes as hyper-ellipsoids in feature space. The third fuzzy classification calculates the Euclidean distance to the 5 nearest reference pixels in each class, making it effectively a fuzzy *k*-nearest neighbour (*k*-NN) algorithm. The algorithm is a non-parametric classification algorithm as it does not make any assumption about the statistical distribution of training pixels (Lucieer, 2006).

In recent years, Support Vector Machine (SVM) classifiers have been introduced and successfully applied in remote sensing research (Pal and Mather, 2005). The main advantage of an SVM classifier is that it is a non-parametric kernel-based classifier. The aim of the SVM classifier is to determine the location of the decision boundaries that produce the optimal separation of classes. The training pixels that lie both between the class centroids and on the edge of the class distributions in feature space (support vectors) are used to define the classification hyperplane. This study applies SVMs with two different kernel types, linear and radial basis. The SVM penalty parameter allows a certain degree of misclassification, controlling the trade-off between allowing training errors (particularly important for non-separable training classes) and forcing rigid margins. The SVM classification output consists of the decision values of each pixel for each class, which can be interpreted as probabilities. This study applies three SVM classifiers: 1) a radial basis kernel with a penalty parameter of 100 and a gamma of 0.25, 2) penalty of 200 and a gamma of 0.5, and 3) a linear kernel with a penalty of 200. In total, seven different soft classifiers were applied.

Ensemble classification

In order to combine the seven different soft classification results an ensemble classifier or multiclassifier system is applied. Several techniques exist to combine classification results, the most commonly methods are voting (committee classifiers), consensus theory, and evidential reasoning (Dempster-Shafer) (Richards and Jia, 2005). In this study we apply a range of combination rules from consensus theory (Benediktsson, 1999; Benediktsson et al., 2007; Briem et al., 2002). Consensus theory involves combining probabilities from multiple experts (in this case classifiers) according to a combination formula, the *consensus rule*. The most commonly used consensus rule is the linear opinion pool (LOP), which computes the joint posterior probability (group probability) of pixel X belonging to class ω_i for S classifiers (Benediktsson, 1999; Richards and Jia, 2005).

$$f(\omega_i|X) = \sum_{s=1}^{S} \alpha_s p(\omega_i|x_s) \tag{2}$$

The source specific weights α_s determine the relative influence of each of the classifiers on the final group probability. One limitation of this rule is that one data source tends to dominate in the decision making. Another acceptable consensus rule that overcomes this limitation is the multiplicative version (LOGP) (Benediktsson, 1999; Richards and Jia, 2005).

$$f(\omega_i|X) = \prod_{s=1}^{S} p(\omega_i|x_s)^{\alpha_s} \tag{3}$$

It is worth noting that zeros are vetoes in this approach, i.e. if one source probability is zero then the group probability is zero as well, irrespective of the recommendations from the other classifiers. The third rule, the logarithmic opinion pool (LOGPL) is a slight variation on Eq. 3

$$\log\{f(\omega_i|X)\} = \sum_{s=1}^{S} \alpha_s \log\{p(\omega_i|x_s)\} \tag{4}$$

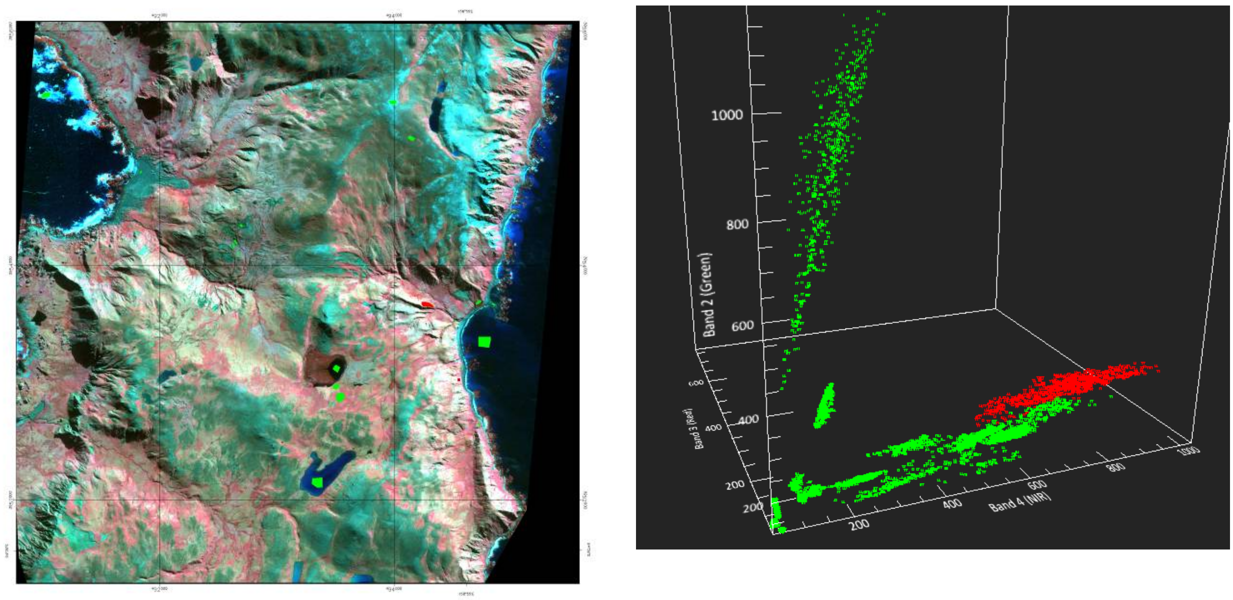
The group probabilities are evaluated to determine the maximum class probability in order to assign hard class labels to each pixel. Two scenarios were run for each consensus rule. Firstly, equal weights were assigned to all classifiers. Secondly, the accuracy of each classification result was assessed and the individual accuracy values were used for the weighting parameter α_s, i.e. the higher the accuracy the higher the contribution of that classifier.

In the final step, we can quantify the disagreement between the classifiers by calculating the normalized Shannon's information entropy. When all classifiers agree (the probabilities and membership values are similar) then the entropy is zero. If the classifiers disagree and their probabilities and membership values are different the entropy is close to one.

$$E_t = \frac{-\sum_{s=1}^{S} p(\omega_t|x_s)\log_2 p(\omega_t|x_s)}{\log_2\left(\frac{1}{S}\right)} \qquad (5)$$

Results

The binary ensemble classification was applied to a subset of the Quickbird image of Macquarie Island (Fig. 1). Fig. 2a shows a false colour composite of the 2.4 m multispectral bands of the study area. The red polygons represent regions of interest (ROIs) for classifier training that were digitized based upon differential GPS (DGPS) field samples. The green ROIs represent the class ‘other’, which includes vegetation communities, bare soil and rock, and water. The separability of the grazed class from the ‘other’ class can be seen in Fig. 2b, which shows a 3D scatter plot of band 4, 3, and 2. A Jeffries-Matusita separability index of 1.957 indicates good separability between the two classes. A separate set of ROIs was digitized from the pansharpened Quickbird bands (0.6 m) based on a combination of DGPS and photo locations and visual interpretation. These independent ROIs were used for accuracy assessment.



a

b

Fig. 2. a) False colour composite of the Quickbird image of the study area on Macquarie Island. The red polygons are training areas for the grazed class and the green polygons are training areas for the ‘other’ class. b) 3D scatter plot (feature space) of the image with the training pixels coloured according to their class, showing class separability in feature space.

Fig. 3 shows three out of seven classification results. Fig. 3a and 3d show the probabilities and hard classes of the maximum likelihood classifier (with the grazed area in red). Fig. 3b and 3e show the classification results for the SFCM with the Euclidean distance metric and Fig. 3c and 3f show the classification results for the SVM classifier with a radial basis function with a penalty parameter of 100 and a gamma of 0.25. The classification accuracies of all classifiers are listed in Table 1. The highest classification accuracy is obtained by the SFCM classifier with the Euclidean distance metric. The lowest individual accuracy is from the SFCM classifier with the Mahalanobis distance metric. The maximum likelihood classifier has a relatively low, but acceptable, accuracy compared to the SVM and SFCM classification results. The accuracies of the ensemble classifiers are very high at 97%, which is higher than most individual classifiers. Only the SFCM classifier with the Euclidean distance metric scores 1% higher. Foody (2007) notes that an ensemble classification need not be more accurate than all of the component classification used in its construction. One very attractive characteristic of the ensemble classifier is that it can provide useful information on classification uncertainty or classifier agreement.

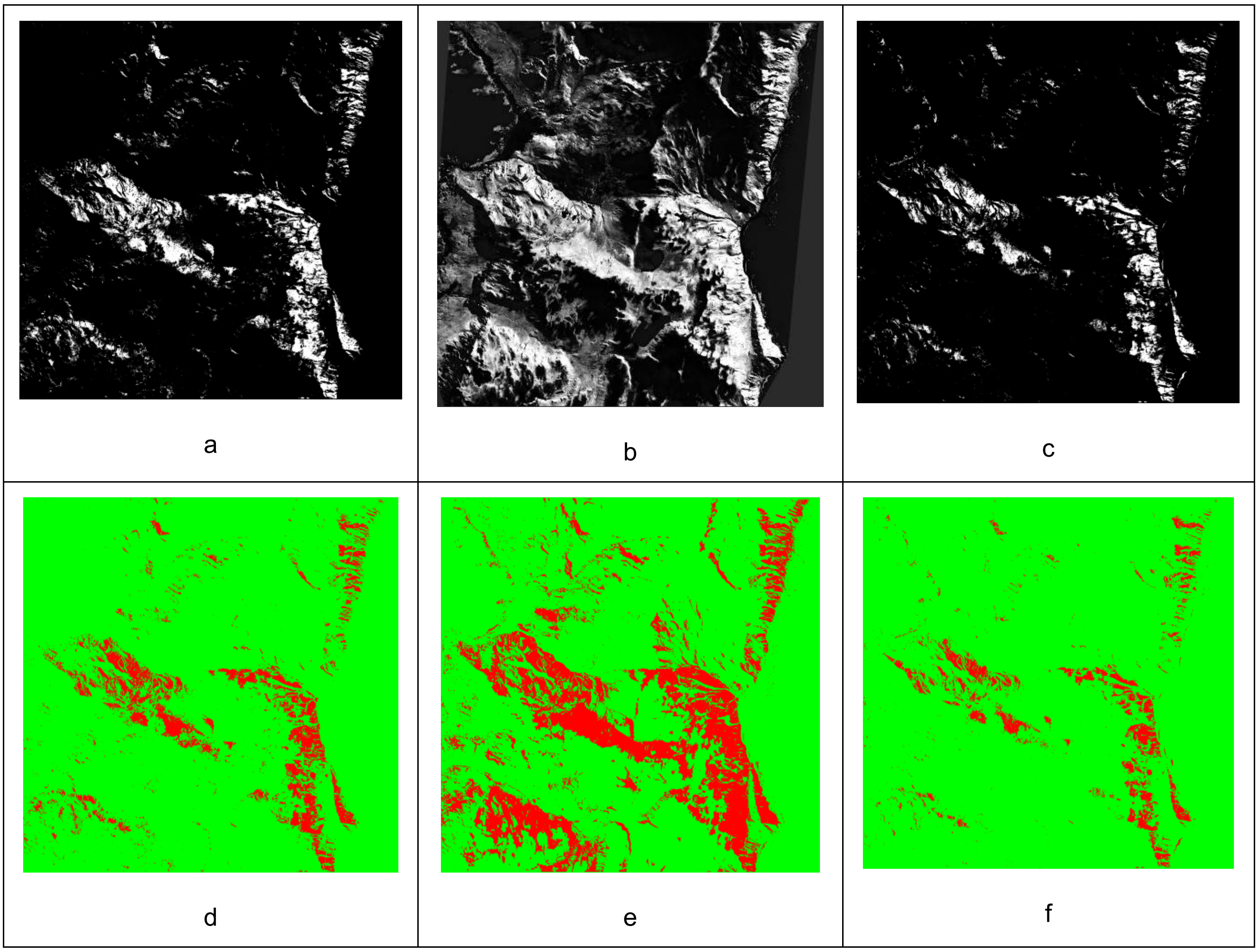
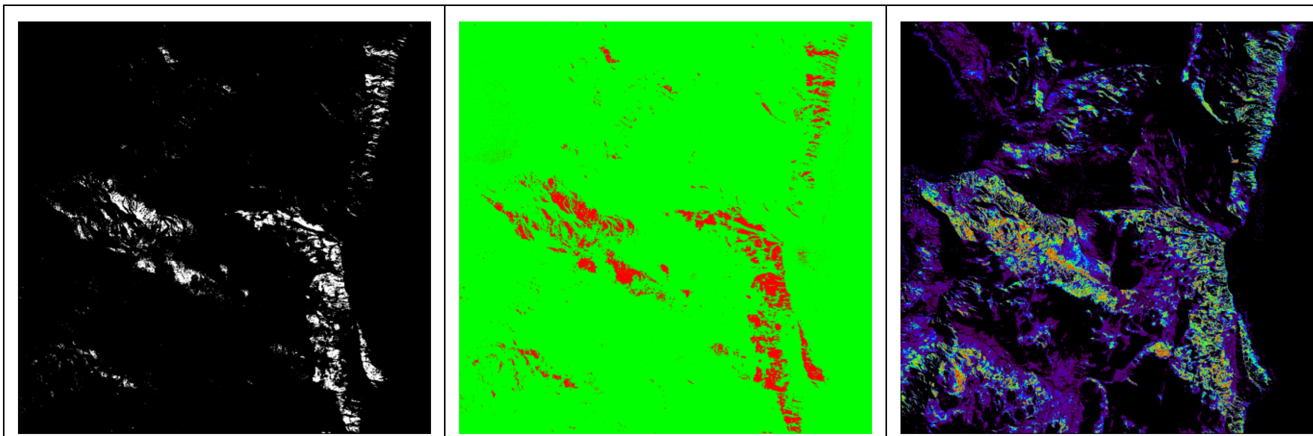


Fig. 3. Three out of seven classification results: a) membership values for the grazed class from the maximum likelihood classifier with the corresponding hard classes in d); b) and e) the SFCM minimum distance classification results; c) and f) the SVM classification results with the radial basis function, penalty of 100, and gamma of 0.25.

Fig. 4 shows the final result of the LOGP (multiplicative consensus rule) classifier weighted according to the individual classifier accuracies. Fig. 4a shows the group probabilities. Fig. 4b shows the hard classes based on the maximum probability. Fig. 4c shows the entropy to highlight classification uncertainty or classifier disagreement. The red areas highlight locations where classifiers disagree. These areas strongly correlate with transition zones (ecotones) where grazed areas slowly change into ungrazed vegetation communities. The overall accuracy of this result is very high at 97.4%. There is very limited variation in the classification results between the different consensus rules. When taking out the two worst performing classifiers (maximum likelihood and SFCM Mahalanobis distance) the classification accuracy does not increase, in fact it drops by 0.3%. The classifier entropy drops significantly when the two worst classifiers are taken out; from a mean entropy of 0.24 for seven classifiers to 0.19 for five classifiers.



a b c

Fig. 4. The binary ensemble classification result for LOGP weighted based on the accuracies of the individual classifiers. a) group probabilities for the grazed class; b) hard classification result based on the maximum probability; c) classifier entropy showing high entropy (in red) where the different classifiers disagree.

Table 1. Classification accuracies for the grazed class.

Classifier	Pixels	Accuracy
Maximum likelihood	5476	0.822
Fuzzy minimum distance	6525	0.980
Fuzzy Mahalanobis distance	3972	0.597
Fuzzy *k*-NN (5 neighbours)	6471	0.972
SVM, Radial, Gamma=0.25, Penalty=100	6448	0.968
SVM, Linear, Penalty=200	6363	0.956
SVM, Radial, Gamma=0.5, Penalty=200	6448	0.968
LOP equal weights	6473	0.972
LOP weighted	6476	0.973
LOGP equal weights	6486	0.974
LOGP weighted	6486	0.974
LOGPL equal weights	6486	0.974
LOGPL weighted	6486	0.974
LOGP weighted 5 best classifiers	6468	0.971
Total grazed reference pixels	6658	

Conclusions

This study applied a binary ensemble classification to identify grazed vegetation communities on sub-Antarctic Macquarie Island from Quickbird imagery. The ensemble classifier was based on seven soft classification algorithms combined with consensus rules. The classification algorithms consisted of the maximum likelihood algorithm, three different supervised fuzzy classifiers with different distance metrics, and three Support Vector Machine classifiers with different parameters. The classification was a binary classification problem, because the focus was on identifying vegetation communities affected by rabbit grazing. The overall accuracy of the ensemble classifier was very high at 97%. The main advantage of the ensemble classifier is that it allows calculation of classifier agreement and uncertainty. The areas of strong disagreement between classifiers corresponded to vegetation transition zones or ecotones. In future work, this approach will be expanded to the whole island and multiple land cover classes.

References

Benediktsson, J.A., 1999. Classification of multisource and hyperspectral data based on decision fusion. IEEE Transactions on Geoscience and Remote Sensing, 37(3 I): 1367-1377.

Benediktsson, J.A., Chanussot, J. and Fauvel, M., 2007. Multiple classifier systems in remote sensing: from basics to recent developments, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 501-512.

Bergstrom, D.M. and Chown, S.L., 1999. Life at the Front: History, Ecology and Change on Southern Ocean Islands. Trends in Ecology and Evolution, 14: 472-477.

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York.

Boyd, D., Sanchez-Hernandez, C. and Foody, G., 2006. Mapping a specific class for priority habitats monitoring from satellite sensor data. International Journal of Remote Sensing, 27(13): 2631-2644.

Briem, G.J., Benediktsson, J.A. and Sveinsson, J.R., 2002. Multiple classifiers applied to multisource remote sensing data. IEEE Transactions on Geoscience and Remote Sensing, 40(10): 2291-2299.

Burrough, P.A., Gaans, P.F.M.v. and MacMillan, R.A., 2000. High-resolution landform classification using fuzzy k-means. Fuzzy Sets and Systems, 113: 37-52.

Copson, G.R. and Whinam, J., 1998. Response of vegetation on subantarctic Macquarie Island to reduced rabbit grazing. Australian Journal of Botany, 46(1): 15-24.

DigitalGlobe, 2008. DigitalGlobe - Quickbird imagery.

Doan, H.T.X. and Foody, G.M., 2007. Increasing soft classification accuracy through the use of an ensemble of classifiers. International Journal of Remote Sensing, 28(20): 4609-4623.

Foody, G.M., 1996. Approaches for the Production and Evaluation of Fuzzy Land Cover Classifications from Remotely Sensed data. International Journal of Remote Sensing, 17(1): 1317-1340.

Foody, G.M., Boyd, D.S. and Sanchez-Hernandez, C., 2007. Mapping a specific class with an ensemble of classifiers. International Journal of Remote Sensing, 28(8): 1733 - 1746.

Kirkpatrick, J.B. and Scott, J.J., 2002. Change in undisturbed Vegetation on the Coastal Slopes of Subantarctic Macquarie Island, 1980-1995. Arctic, Antarctic, and Alpine Research, 34: 300-307.

Lu, D. and Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. International Journal of Remote Sensing, 28(5): 823 - 870.

Lucieer, A., 2006. Fuzzy classification of sub-Antarctic vegetation on Heard Island based on high-resolution satellite imagery, IEEE International Geoscience and Remote Sensing Symposium (IGARSS'06). IEEE, Denver, Colorado, USA.

Pal, M. and Mather, P.M., 2005. Support vector machines for classification in remote sensing. International Journal of Remote Sensing, 26(5): 1007-1011.

Richards, J.A. and Jia, X., 2005. Remote Sensing Digital Image Analysis - An Introduction. Springer, Berlin.

Tso, B. and Mather, P.M., 2001. Classification Methods for Remotely Sensed Data. Taylor and Francis.

Tzeng, Y.C., Chiu, S.H., Chen, D. and Chen, K.S., 2008. Multisource remote sensing images classification/ data fusion using a multiple classifiers system weighted by a neural decision maker, International Geoscience and Remote Sensing Symposium (IGARSS), pp. 3069-3073.

Zadeh, L.A., 1965. Fuzzy Sets. Information and Control, 8: 338-353.

Zhang, J. and Foody, G.M., 2001. Fully-fuzzy Supervised Classification of Sub-urban land Cover from Remotely Sensed Imagery: Statistical and Artificial Neural Network Approaches. International Journal of Remote Sensing, 22(4): 615-628.